


Building a Dataset of Pointing Gestures for Elderly People in Iraqi Nursing Homes

Kawther Thabt Saleh^{1,2*} , Abdulmir Abdullah Karim³ 

¹ Informatics Institute for Postgraduate Studies, Iraqi Commission for Computers & Informatics, Baghdad 10071, Iraq

² Department of Computer Science, College of Education, Mustansiriyah University, Baghdad 10052, Iraq

³ Department of Computer Science, University of Technology, Baghdad 10066, Iraq

Corresponding Author Email: phd202130684@iips.edu.iq



Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290632>

ABSTRACT

Received: 30 March 2024

Revised: 23 May 2024

Accepted: 22 July 2024

Available online: 25 December 2024

Keywords:

gesture, elder, stroke patients, bounding box annotation, semantic labelling, augmentation, YOLOv8

Elderly and stroke patients with speech and hearing difficulties often struggle with communication, particularly in complex environments. This paper presents a new dataset specifically created for semantic segmentation, instance segmentation, and sign recognition to meet the needs of these individuals. The dataset pertains to elderly Iraqis and was collected from several nursing homes. Approval from the relevant authorities has been obtained. The images in the dataset have been pre-processed, enhanced, and annotated to facilitate effective model training and evaluation. By applying the YOLOv8 model for gesture detection and classification, we achieved precision, recall, and mAP50 scores of 95.6%, 93.9%, and 97.1%, respectively. These findings highlight the dataset's potential to improve assistive technologies for better communication support.

1. INTRODUCTION

Elderly people and stroke patients often have communication difficulties due to hearing and speech limitations. Doctors find it challenging to communicate with stroke patients who have trouble speaking. Caregivers and those who look after the elderly also face difficulties communicating with them. Additionally, there is currently no dataset for gestures specifically for elderly stroke patients in Iraq; only datasets for well-known sign languages for the deaf and mute are available. This research aims to develop a comprehensive solution for improving communication among elderly individuals and stroke patients, their caregivers, and those around them.

This research aims to serve the elderly and stroke patients and aid those who take care of them. Moreover, it holds significant potential for societal impact by improving communication for elderly and stroke patients. Through innovative methods, the study seeks to enhance the quality of life for these vulnerable populations.

In these situations, nonverbal communication through gestures is essential for promoting contact [1]. Body language is a crucial non-verbal communication tool that aids in comprehending emotions, intentions, and messages across various social and cultural contexts [2]. Hand gestures enable the nonverbal expression of emotions [3], which is particularly beneficial for stroke patients and older adults struggling with communication, thereby reducing frustration, and enhancing social interaction. Hand gestures significantly enhance social interaction and interpersonal communication by expressing interest, engagement, empathy, and understanding, thereby strengthening moral relationships despite communication gaps [4]. Computer interaction technology is considered a modern

means that contributes to improving the quality of life of individuals who face communication difficulties due to speech and hearing limitations, in addition to individuals affected by health problems such as strokes. We propose a large dataset aimed at sign gesture identification, emphasizing instance and semantic segmentation, to help developments in assistive technology. The dataset has been carefully selected to help researchers create reliable models that can improve communication for this category of people. By utilizing innovative methods like YOLOv8n, this study aims to improve the quality of life for elderly individuals and stroke patients, who are considered vulnerable populations. we used Yolov8 to detect and classify the sign gesture motions for meet patients and elder's needs who have troubles in speaking or hearing. This is done by training the YOLOv8 model on the constructed data. Advanced image processing and feature extraction techniques were used to efficiently the model's execution in accurate results.

Moreover, the confusion matrix, F1 confidence curve, precision curve, recall curve, and precision-recall curve are used for evaluating the model's performance in accurately identifying and classifying different sign gesture motions.

The rest of the paper is organized as follows: Section 2 reports the survey of the literature. In Section 3, we discuss the proposed methodology. In Section 4, we explain Results and Discussion in Section 5, we discuss the method of augmentation used in preparing the dataset. Finally, Section 8 contains the conclusion reached in this study.

2. RELATED WORK

There has been an increase in non-verbal communication

research during the last few decades. To comprehend these communications, specialists and scholars have created databases. This study examines and evaluates earlier studies on gesture databases, showcasing the work of academics and enhancing our knowledge of this vital element of interpersonal communication.

In 2023, Alsulaiman et al. [5] analyzed 17 Arabic sign language datasets in 2023 and proposed a strategy for building a sign language dataset. They used the King Saud University Saudi-SSL (KSU-SSL database), which has 293 signs and 33 signers and introduced a 3-convolution graph neural network architecture for sign language recognition.

Shin et al. [6] developed a multi-branch network for sign language recognition in 2023, utilizing local feature analysis and transformer computation. The model achieved 98.30% accuracy in the lab dataset and 89.00% accuracy in the KSL dataset with 77 labels.

In June 2023, Johari et al. [7] made available the MyWSL2023 (MyWSL) dataset. It has 3,500 images of 10 static Malaysian Sign Language situations, gathered from 5 participants aged 20-21.

In 2022, Cassim et al. [8] gathered hand gestures data using numerous classification methods, including k-nearest neighbors, neural networks (NN), support-vector machines (SVM), decision trees (DT), and random forests (RF), after gesture execution.

Zhou et al.'s [9] 2022 framework regain key features from gesture data, providing effective data for the bidirectional long short-term memory (BLSTM) model for CSLR. They examined the framework of a difficult Hong Kong sign language (HKSL) dataset, studying hand gesture data from 6 signers for 50 sentences.

Kasapbaşı et al. [10] developed a 2022 dataset to translate hand motions and labels in the American Sign Language alphabet (ASLA) using 104,000 photos. The dataset, aided by a convolutional neural network, achieved 99.38% accuracy with minimal loss, despite the dataset's volume and various conditions.

Suharjito et al. [11] used a transfer learning approach and a Convolutional Neural Network model to solve Sistem Isyarat

Bahasa Indonesia (SIBI) dataset issues in 2021. They collected 200 movies. The model with the most frozen inception modules achieved the greatest validation accuracy was attained at 100% and the best testing accuracy was at 97.50% after training and testing.

Singh [12] developed a 3D convolutional-based neural network in 2021 to simulate hand motions in the Indian community, providing natural language outputs corresponding to standard Indian Sign Language (ISL) signs. The model can help with speaking with dumb and deaf individuals and be applied in various fields, including medical and industrial.

In 2021, Zhang et al. [13] used 5892 reliable videos from 30 deaf-mutes in 2021 for training and testing. They selected data from 27 individuals in the training set and 595 videos in the test set, using enhanced FlickrNet for spatiotemporal information extraction.

In 2020, Adithya and Rajesh [14] presented a video dataset of hand gestures used in emergency situations to represent words in Indian Sign Language (ISL). The dataset included eight ISL phrases from 26 people aged 22-26, recorded in an interior space with typical lighting. ISL is classified using support vector machines and deep learning.

In 2020, Pacifici et al. [15] released EMG and IMU data from 26 Italian Sign Language alphabet movements using the Myo Gesture Control Armband. The dataset included 780 samples, 30 captures per move, and was gathered from a 24-year-old individual in a lab environment at 200 Hz for 2 seconds.

In 2020, Wadhawan and Kumar [16] developed a deep learning-based model for static sign language identification using convolutional neural networks. They analysed 35,000 sign images from 100 users and found that the training accuracy of their method was 99.72% for colourful images and 99.90% for grayscale images.

In 2019, Latif et al. [17] used machine learning deep learning, and, computer vision algorithms to create automated solutions for the deaf and hard of hearing using the Arabic Sign Language (ArSL) dataset. The ArSL2018 dataset includes 54,049 photos from 40 participants.

Table 1. Comparison between earlier research that has constructed and classified words\sentences sign language datasets

Work	Year	Constructed Dataset	Dataset Description	Recognition/Classification Method	Accuracy%
Alsulaiman et al. [5]	2023	KSU-SSL	Samples: 145,035, signs: 293, and signers: 33.	CGCN	97.25 average ACC
Shin et al. [6]	2023	KSL	Videos: 21, sample images: 20 and Korean daily activities: 77	modified version of FFN of the ViT	98.30
Johari et al. [7]	2023	MyWSL2023	Images: 3,500, signs: ten, participants: 5(men: 2 and women: 3) and ages: 21 - 20.	simple Sequential CNN	98
Zhou et al. [9]	2022	HKSL	Signers: 6 and sentences: 5.	BLSTM	9.4
Suharjito et al. [11]	2021	SIBI	200 videos by two signers for ten words (classes).	CNN	testing ACC: 97.50%, validation ACC: 100%.
Singh [12]	2021	ISL	participants: 10, gestures: 20	3D convolution-based CNN	training ACC: 99.67% & validation ACC: 88%.
Ko et al. [18]	2019	KETI	Videos: 14,672.	Open Pose	validation ACC: 93.28% and test ACC 55.28%.
Our (SGESP) Dataset	2024	SGESP	Images: 2,860, signs 26 and participants: 110.	Yolov8	Precision: 0.956, recall: 0.939, mAP50: 0.971, and mAP50-95: 0.775.

Table 2. Comparison of previous works that have constructed and classified letters\numbers of sign language datasets

Work	Year	Constructed Dataset	Dataset Description	Recognition/Classification Method	Accuracy%
Cassim et al. [8]	2022	American Sign Language fingerspelling.	Sessions: 10 recording, each session: 15 recordings for every gesture, and classes: 27.	k-nearest neighbours, SVM, NN, decision trees, and random forests	85.51
Kasapbaşı et al. [10]	2021	ASLA	The dataset was collected under varying lighting conditions at different times of the day.	CNN	99.38
Latif et al. [17]	2019	ArSL2018	Arabic Signs: 32, images: 54,049 and participants: 40.	\	\

In 2019, Ko et al. [18] introduced the sign language KETI dataset, consisting of 14,672 videos. The dataset serves Korean sign language translation. They developed a neural network model using human key points extracted from hands, face, and other body parts. The model achieved 93.28% and 55.28% translation validation accuracy and test set for 105 sentences, making it suitable for emergency situations. The researchers compared various neural sign translation models using classical metrics to measure their performance. Table 1 presents a comparison between our constructed dataset for classifying gestures for the elderly and patients with strokes who are unable to speak, and earlier research that has constructed and classified sign language datasets. It's important to note that all of these datasets typically involve gestures, words, and often phrases, as they aim to facilitate communication through visual means rather than spoken language. Table 2 presents a comparison of some previous works that have constructed datasets for sign language and

utilized letters and numbers. All the datasets listed below pertain to sign language for the deaf and mute. These sign languages are taught from a young age and have specific dictionaries for each country. However, the datasets we constructed are intended for elderly individuals and stroke patients who have lost their hearing and speech later in life. These patients find traditional sign language dictionaries too complicated to use. Our gestures were developed and validated after consulting several nursing homes and doctors who work with stroke patients, ensuring the gestures are practical and approved for their use.

3. PROPOSED METHODOLOGY

The suggested system diagram is shown in Figure 1. The stages methodology is illustrated Figure 1.

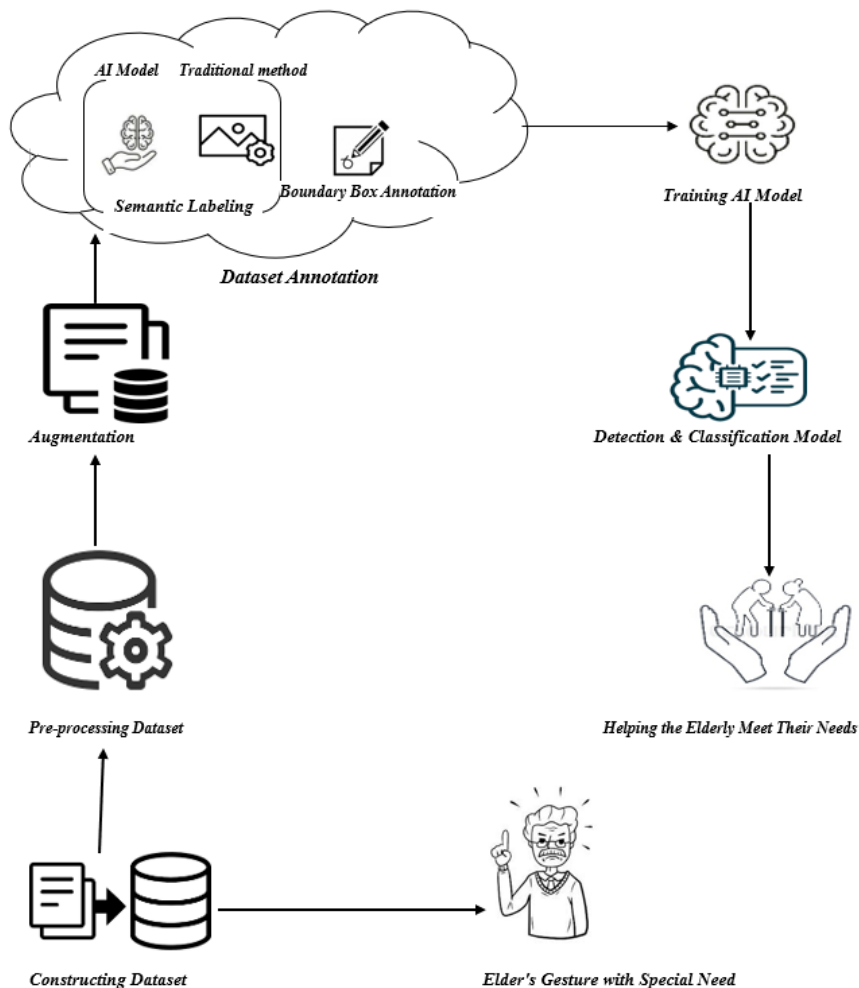


Figure 1. Methodology of the proposed system

3.1 Dataset collection and construction









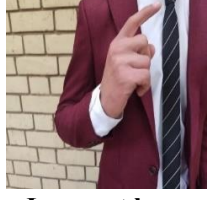

















The database's images were carefully captured to provide an accurate and diverse its representation of the fundamental hand gestures that elderly people and patients with strokes are using in their daily life needs. Here's how to collect images explained:

- Dataset approved: The dataset was validated through

visits to multiple nursing homes and consultations with doctors who work with stroke patients, ensuring that the gestures are practical and suitable for meeting the daily needs of patients.

- Classes selected: A total of 26 classes were chosen to represent the fundamental hand motions approved by visiting several nursing homes for the elderly to meet their daily needs. The total number of datasets for the original images is 2860.

Table 3. SGESP dataset for the hand gesture of the elderly and stroke patients who suffer from hearing and speech impairment with its meaning

			
Hello	I am good	Thank you	I want take a shower
			
Wait	Correct	I want to eat	Finished
			
I can not hear	I want to change my clothes	I want to drink water	Me
			
You	Upset	Come	I am not sure
			
I am married	What	Why	Excuse me
			
I love you	Please	Listen	Stop
			
It is Time	Help me		

- Captured dataset images: Different webcam and smartphone lenses were used for each class to guarantee a range of lighting and image quality. Web camera (Logitech 1080p) and the mobile camera (The Poco X3 Pro phone features a quadruple rear camera with 48 megapixels, dual PDAF image stabilization, 8-megapixel resolution, 2-megapixel macro, and depth sensor. It supports flash, panorama, HDR, and high-quality video. The front camera has a 20-megapixel wide angle) were used to capture the dataset images.

- Selection of participants: A culturally and physically diverse group of individuals of different ages and genders, including patients with strokes and the elderly, had been selected. There are 110 samples from various participants in each class.

- Create multiple circumstances: To offer diversity in data and show various situations (indoor and outdoor) that users may meet in daily life, images are taken with varying backgrounds, lighting conditions, and orientations.

- Different environments and complex backgrounds: The gestures were captured in diverse settings, including complex backgrounds and environments with colors approximating skin tones, to improve the model's performance during both training and testing phases.

To satisfy elderly daily needs, a scalable dataset constructing the variety of hand gestures used by stroke patients and the elderly has been established. This makes it easier to use the dataset for research and development in fields like assistive technology and machine learning. This constructed dataset is named "Sign Gestures for Elders and Stroke patients having Limited Hearing and Speech Dataset" with abbreviated name (SGESP). Table 3 presents part of the data set for the hand gesture of the elderly and stroke patients

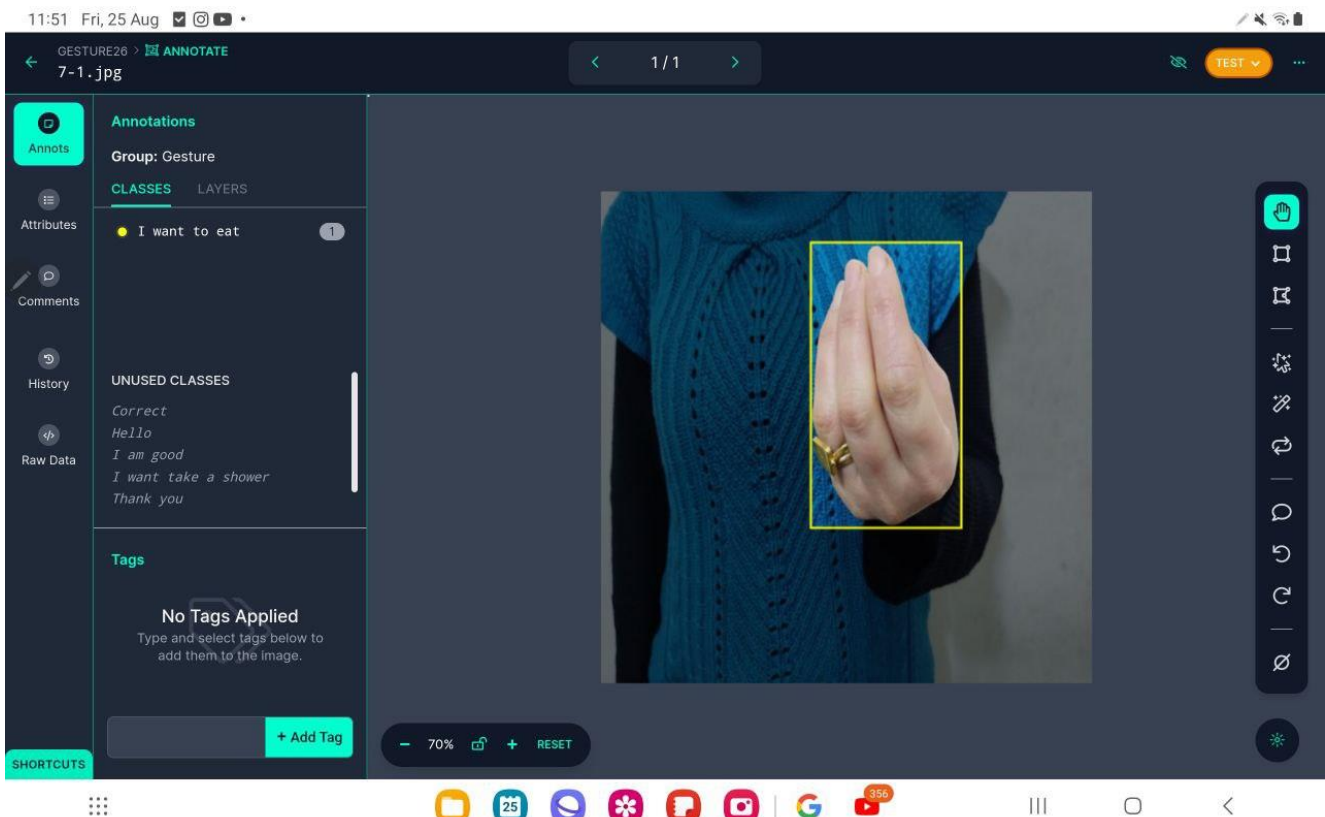
who suffer from hearing and speech impairment with its meaning.

3.2 Preprocessing and annotation

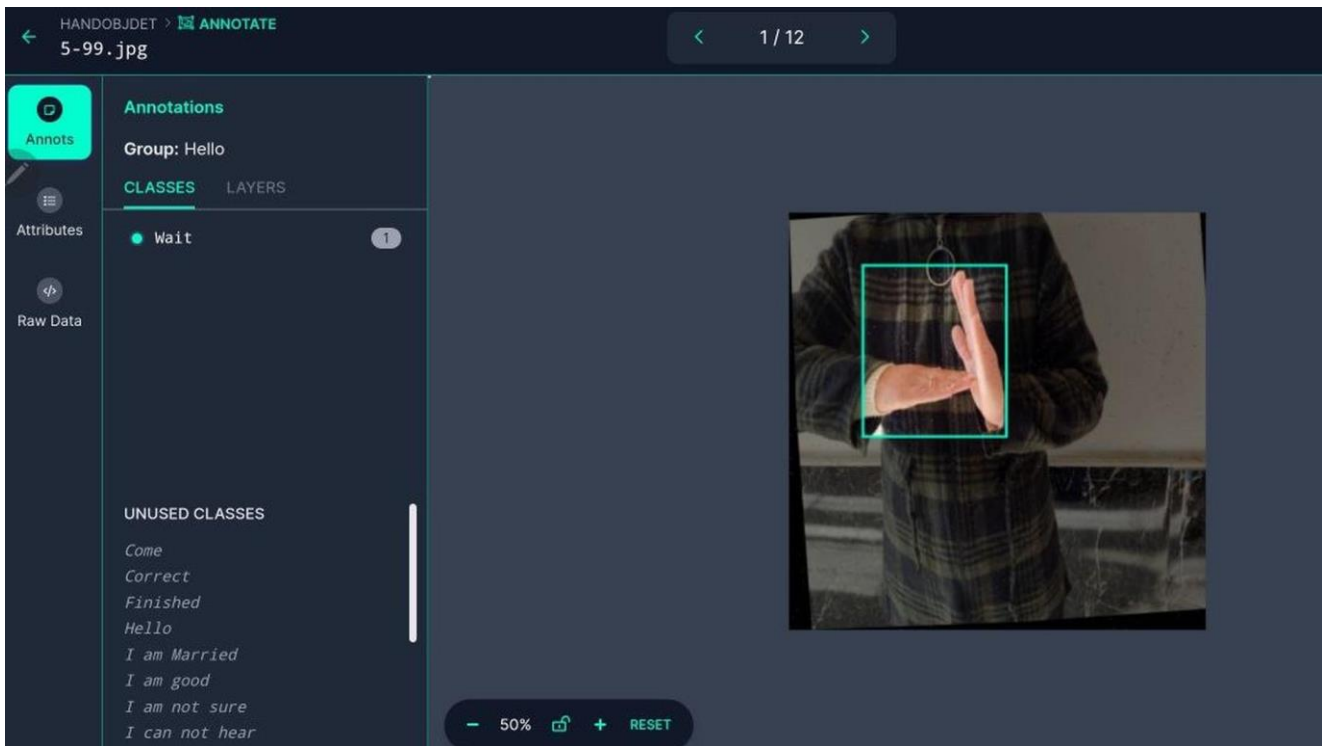
As a preprocessing step, cropping and resizing processes were performed. Manual cropping was used to remove areas that were uninteresting in the image, which helped focus the main content. Antialiasing technology was also used during the resizing process. These visual processes enhance the effectiveness of pre-processing to improve the readiness of images for use in visual analysis or the training of deep learning models. All the dataset images were resized to 640 × 640 pixels for reliable model training. Convert the image type in the dataset to JPG. Each image was annotated with instance-level and semantic-level annotations. While semantic annotations categorize pixels into pertinent gesture categories, instance annotations identify specific gesture-related areas in particular pixels.

3.2.1 Bounding box annotation

By explicitly defining items of interest, the core approach of bounding box annotation gives visual data structure. It enables machine learning models to grasp item positions, sizes, and spatial connections by making training and assessment of those models easier. The object recognition, tracking, and segmentation capabilities of this annotation approach enable a wide range of applications, promoting improvements in autonomous cars, medical imaging, surveillance, and more. Bounding box annotation fills the semantic gap between unprocessed photos and computer vision, enabling accurate and reliable AI systems to perceive and efficiently communicate in the visual world [19, 20].



(a) "I want to eat" image annotation



(b) "Wait" image annotation

Figure 2. The Roboflow tool for bounding box annotation examples

Bounding Box Annotation may be done with a variety of tools. Roboflow is an effective Bounding Box Annotation tool [21]. We employed Roboflow to analyze hand gesture images which were generated by elderly or stroke patients who had difficulty speaking. Bounding boxes are drawn around the hand gestures in the dataset in order to train deep learning models to detect and classify gestures. As a result, the outcomes are more reliable and accurate. This accurate process ensures that each hand gesture in the image gets a suitable description of its spatial extent. Figure 2 shows examples of the bounding box annotation using the Roboflow tool. Table 4 provides examples of a sample of images along with their corresponding masks, demonstrating the segmentation process.

Table 4. Examples of a sample of images and their masks

Original Image	Mask Image	Original Image	Mask Image
Original Image	Mask Image	Original Image	Mask Image
Original Image	Mask Image	Original Image	Mask Image

3.2.2 Semantic labelling

"Semantic Labelling" or "Pixel-wise Annotation" is the process of assigning a precise label that accurately describes the object or class to which each pixel in an image belongs. As a result, a label that accurately describes each region or item in the image has been assigned. Semantic label is used for Improve recognition accuracy, accurately understand images, Robotics and autonomous vehicle applications, Land and Environment Analysis, Pathology Applications [22, 23].

Without requiring any additional software, ImgLab is a free platform for image annotation that can be used directly by the website [24], which has been used in this work to semantic label annotation for annotate gestures in some of the images to analyze the images that were generated from strokes patients and elderly people who have difficulty speaking and converting them into a mask:

Stage 1: Segmentation by Thresholding

The entire pre-processed dataset was segmented using thresholding.

About 10% of the images were accurately segmented.

Stage 2: K-means Clustering

K-means clustering was applied to the remaining 90% of the dataset images.

About 30% of the resulting images were accurately segmented.

Stage 3: Using Imglab Tool

The Imglab tool was used on the remaining 60% of the images.

A JSON file was generated and converted into segmented images.

About 10% of the resulting images were accurately segmented.

Results Analysis:

Approximately 50% of the images were accurately segmented.

These accurately segmented images were used to train a U-

net model to obtain the best weights.

The remaining 50% of images, which were not accurately segmented, were used for testing.

Final Result:

The U-net model achieved 100% accuracy in segmenting the images.

Table 3 show some examples of images with their masks.

3.3 Augmentation

Augmentation methods are used to improve datasets efficiency by increasing the dataset's variety. These methods include flips, rotations, translations, and brightness changes [25, 26]. By using Augmentation, we simulate differences in gesture representation that occur in real life. Also, to improve the effectiveness of model training and obtain better results, we employ dataset augmentation. For the instance segmentation dataset, the original dataset image $\times 4$ by horizontal flipping, 5% rotate, 3% blurring, and 2% noising is applied on training dataset. The instance segmentation dataset is 10.010 for train (92%), 512 for valid (5%), 286 for test (3%), (10.808) the total size of image and (10.808) text file as annotation label. For the semantic segmentation dataset, the original dataset image $\times 3$ by horizontal flipping, 45% rotate, 9% scale is done on training dataset. Train dataset is 10,400 (number of training samples is 9356 and number of validation samples is 1040) and test dataset is 260.

3.4 Gesture classification

We used the YOLOv8n model for gesture detection and categorization as proof of concept. The created gesture dataset was used to train the model, and it showed promise in successfully identifying and categorizing motions. The dataset's usefulness in creating strong gesture recognition models is demonstrated by the YOLOv8 architecture.

3.4.1 Yolov8

In January 2023, Ultralytics, the company behind YOLOv5, introduced YOLOv8, the latest version of its object detection

model, on GitHub [27]. YOLOv8 introduces significant advancements, including an anchor-free detection head that predicts an object's center directly, eliminating the need for predefined anchor boxes. Additionally, a novel loss function is proposed to improve accuracy. The architecture of YOLOv8 consists of two main components: the backbone for feature extraction and the detection head for object prediction [28]. The backbone comprises convolutional layers that process images at multiple resolutions, enhancing feature extraction efficiency. Notably, a (3×3) convolution now replaces the previous (6×6) convolution in the stem, and the C2f module has replaced the C3 module used in YOLOv5, improving feature concatenation and overall performance. Unlike YOLOv5, where channels must match for concatenation, YOLOv8 performs direct concatenation, optimizing computational efficiency. These architectural modifications—combined with the lightweight design—enable YOLOv8 to achieve improved detection accuracy while maintaining computational cost efficiency [29]. The updated design is illustrated in Figure 3.

The first convolution's kernel was modified from (1×1) to (3×3) in the neck. This reduces the number of parameters and the tensors' total size [23]. YOLOv8 is regarded as being extremely efficient and is compatible with a wide range of hardware configurations, including single and multiple GPUs. Additionally, YOLOv8 offers several model sizes for segmentation, classification, and detection. Larger models are slower than tiny ones since they require more calculations to create a deeper network that can yield more accurate results. The detection models with their mean average precisions, number of parameters, and number of Floating-Point Operations (FLOPs) are shown in Table 5 [30]. Due to their superior detection speed and accuracy, YOLO family algorithms have been extensively developed for real-world applications [31]. YOLOv8, in particular, offers several advantages for detection and classification, including enhanced speed and accuracy [32], free anchoring [33], the ability to focus on different areas of an image [34], high scalability [35], and real-time processing capabilities [36]. For these reasons, we chose to use YOLOv8 in our work.

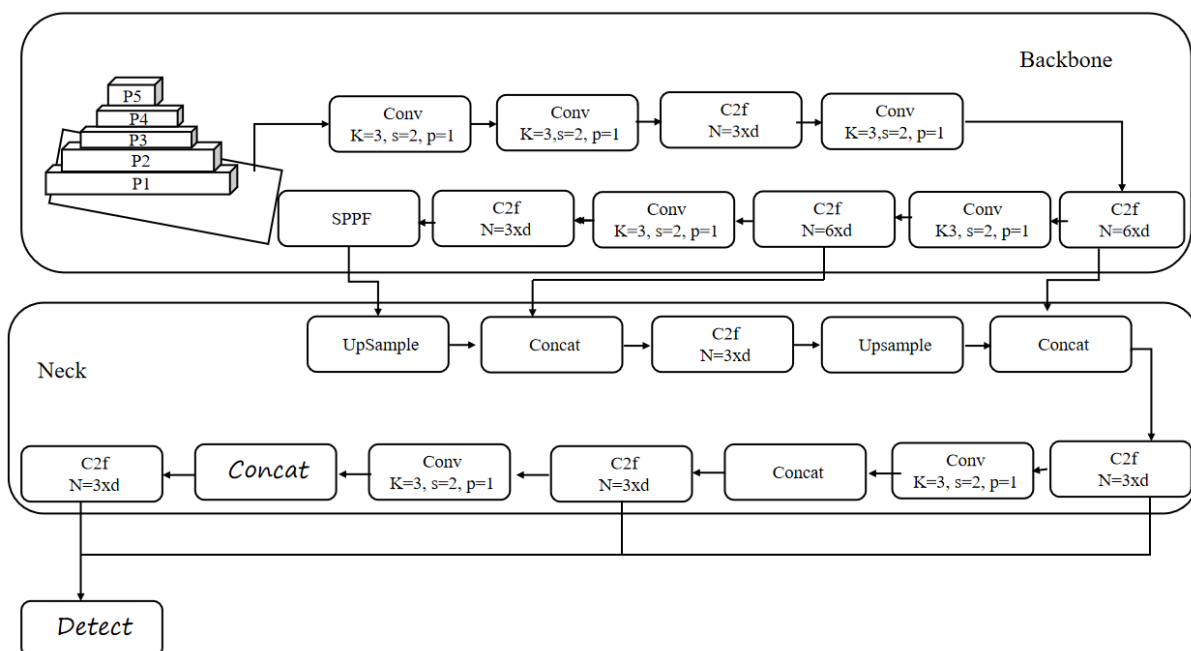


Figure 3. YOLOv8 model structure [29]

Table 5. YOLOv8 detection models [24]

Model	Size	Layers	mAP	Parameters	FLOPs
YOLOv8n	Nano	168	37.3	3.2 million	8.7 billion
YOLOv8s	Small	168	44.9	11.2 million	28.6 billion
YOLOv8m	Medium	218	50.2	25.9 million	78.9 billion
YOLOv8l	Large	268	52.9	43.7 million	165.2 billion
YOLOv8x	Extra Large	268	53.9	68.2 million	257.8 billion

4. RESULTS AND DISCUSSION

The created sign gesture dataset has enormous potential for assisting in the development of assistive models for stroke patients and the elderly who have hearing and speech difficulties. The information this dataset conveys can help develop creative communication solutions that improve their quality of life by enabling precise gesture segmentation and recognition. To detect and classify gestures, we utilized the YOLOv8n model. There were 3010718 total parameters utilized. When we evaluated the classification results, we

found that the precision, recall, mAP50, and mAP50-95 values were respectively 0.956%, 0.939%, 0.971%, and 0.775%. The training process and hyperparameters of the YOLOv8 model are presented in Table 6. The evaluation matrices for each class are illustrated in Table 7 where Box(p) is precision and R is recall and confusion matrix evaluation metric illustrated. It is noted that the Precision value for the Hello class was low compared to the rest of the classes, as it was 0.648. The reason is that the "Hello" sign is similar to the "Finished" sign, which leads to confusion in the classification process. Furthermore, it can be observed that the Why class had a lower recall value (0.55) than the other classes. The reason is that the "Why" sign is similar to the "What" sign which makes confusion in the categorization result. We also noticed that the values of mAP 50-95 were somewhat low compared to the values of mAP 50, as they are limited to the range (0.668-0.863), and the reason for this decrease is over different IoU thresholds, from 0.5 to 0.95 in Figure 4. In order to assess the performance of the model for each class, evaluation matrices such as precision, recall, F1-score, and accuracy were computed. These metrics provide detailed insights into the performance of the model across different categories. The evaluation results are summarized in Table 7, where the performance of each class is thoroughly analysed.

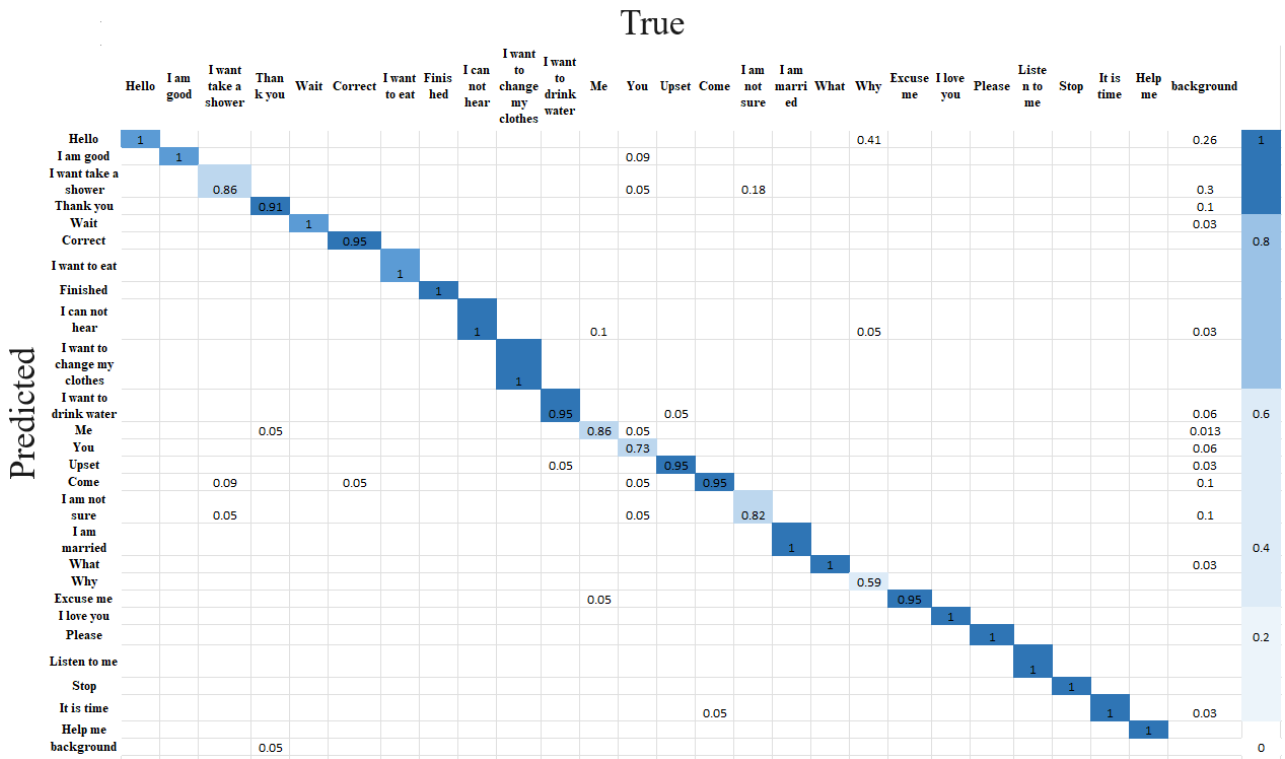


Figure 4. Confusion matrix evaluation results

Table 6. Training process and hyperparameters of the YOLOv8 model

Epochs Completed	196 epochs
Training Duration	2.680 hours
Speed per Image	Preprocess: 0.4ms, Inference: 0.6ms, Loss: 0.0ms & Postprocess: 0.8ms
Model Summary	Layers: 225, Parameters: 3015918, Gradients: 3015902, GFLOPs: 8.2
Transfer Learning	Transferred 319/355 items from pretrained weights
Optimizer	SGD optimizer with learning rate (lr) set to 0.01, and parameter groups defined for weight decay
Training Dataset	5994 images
Validation Dataset	568 images
Early Stopping	Training stopped early due to no improvement observed in the last 50 epochs
Best Model	Epoch 146 was identified as the best performing epoch, and the best model was saved

Table 7. The evaluate matrices for each class

Class	Images	Instances	Box(p)	R	mAP50	mAP50-95
all	567	567	0.956	0.939	0.971	0.774
Hello	567	22	0.648	0.955	0.941	0.798
I am good	567	22	0.958	1	0.995	0.793
I want take a shower	567	22	0.829	0.818	0.891	0.723
Thank you	567	22	0.952	0.901	0.951	0.695
Wait	567	22	0.991	1	0.995	0.812
Correct	567	22	0.999	0.955	0.986	0.798
I want to eat	567	22	0.991	1	0.995	0.8
Finished	567	21	1	0.965	0.995	0.808
I cannot hear	567	21	0.909	1	0.989	0.746
I want to change my clothes	567	22	0.997	1	0.995	0.799
I want to drink water	567	21	0.969	0.952	0.959	0.748
Me	567	21	0.987	0.905	0.958	0.703
You	567	22	1	0.732	0.932	0.75
Upset	567	22	0.955	0.968	0.979	0.825
Come	567	22	0.861	0.955	0.979	0.786
I am not sure	567	22	0.904	0.857	0.933	0.752
I am married	567	22	1	0.969	0.995	0.785
What	567	22	1	0.965	0.995	0.703
Why	567	22	1	0.55	0.822	0.668
Excuse me	567	21	0.953	0.957	0.993	0.803
I love you	567	22	0.993	1	0.995	0.816
Please	567	22	0.994	1	0.995	0.788
Listen to me	567	22	0.994	1	0.995	0.803
Stop	567	22	0.993	1	0.995	0.863
It is time	567	22	0.991	1	0.995	0.731
Help me	567	22	0.993	1	0.995	0.829

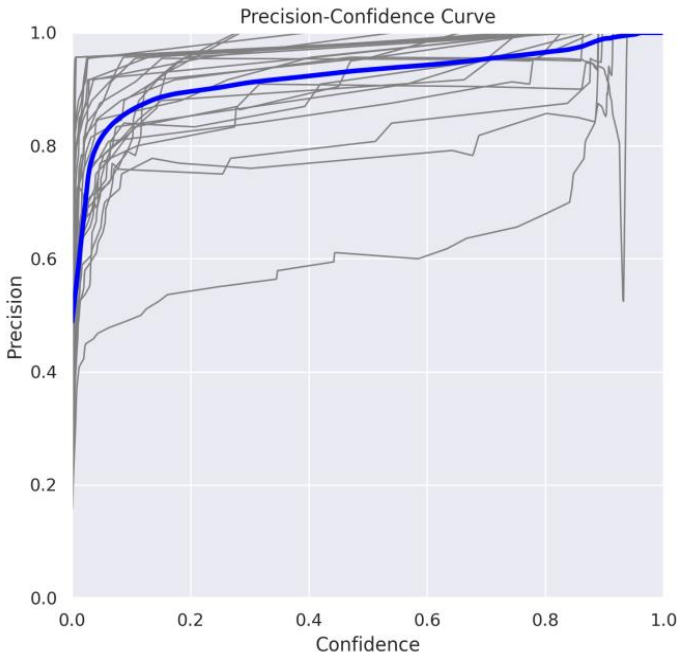


Figure 5. Precision confidence curve

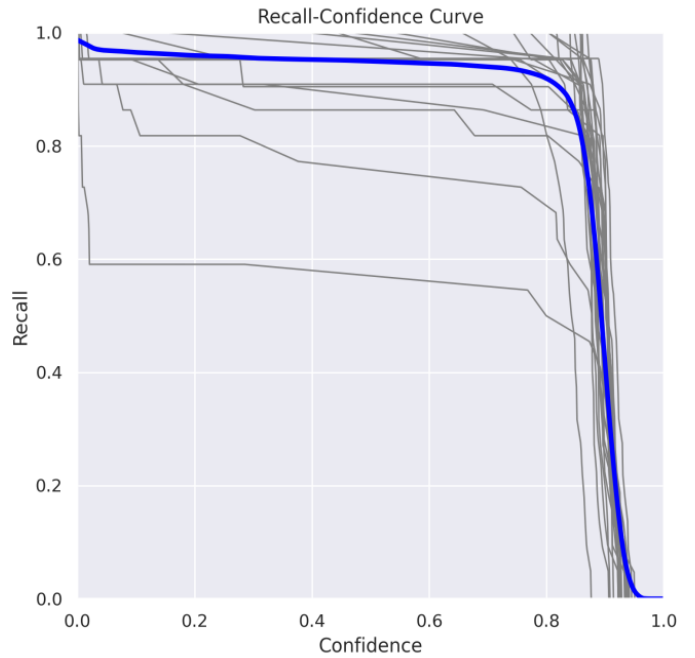


Figure 6. Recall confidence curve

Figure 5 shows that the accuracy of the model generally increases as its confidence level increases. At a confidence level of 96.3%, the accuracy of the model is 100%. At a confidence level of 100%, the accuracy of the model is 80%.

The Figure 6 Illustrated “true confidence curve” shows that the deep learning model is highly accurate, with 90% of images correctly classified at a confidence level of 0.8. The True Confidence Curve shows that the model is very confident in its classifications, with confidence greater than 0.8 for most

images. The “Recall Score” shows that the model is able to distinguish different classes well, as the recall rate varies greatly between different classes.

Figure 7 shows the YOLO F8 model with a high level of confidence in its classification of objects. The optimal operating point indicates that the model can achieve a good balance between confidence level and prediction accuracy. The confidence level of the model is high (0.94) and the prediction value is high (0.728).

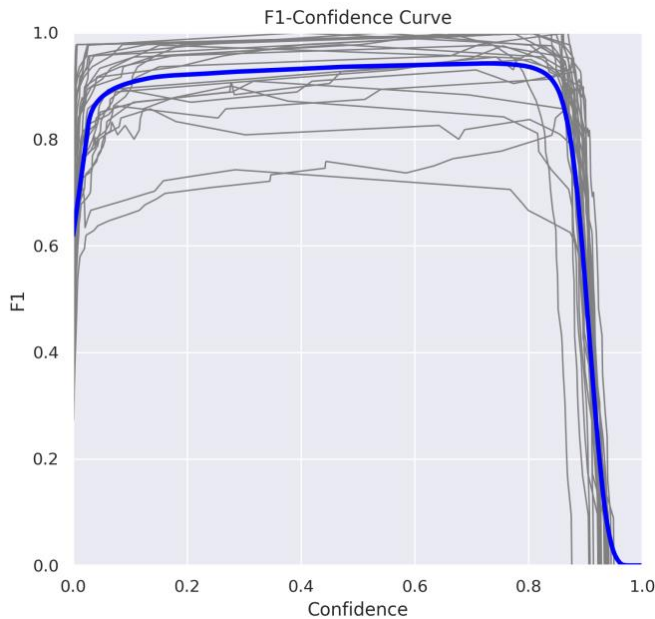


Figure 7. F1 confidence curve

3. CONCLUSIONS AND FUTURE SCOPE

In conclusion, the Sign Gestures Dataset for Elders and Stroke Patients (SGESP) represents a significant advancement in overcoming communication barriers for individuals with limited hearing and speech abilities. Through instance and semantic segmentation, and gesture categorization using YOLOv8 architecture, SGESP not only pushes the boundaries of computer vision but also has profound implications for assistive technology.

Accurate annotation and semantic labelling enhance SGESP's applicability beyond gesture recognition, enabling applications in automatic classification, identity verification, and other fields. Augmentation techniques ensure robustness and adaptability, crucial for addressing real-world scenarios faced by elderly and stroke patients, thereby enhancing scalability and effectiveness.

Looking ahead, SGESP holds promise in fostering greater engagement and communication for elderly people and stroke patients, promoting independence, and enriching daily lives. Future research leveraging SGESP could include expanding the dataset to encompass a broader range of gestures and scenarios, exploring alternative model architectures to improve accuracy and efficiency, and developing real-time gesture recognition systems for immediate interaction needs.

The continued development and application of SGESP are expected to drive further innovations in assistive technologies, ultimately enhancing accessibility and inclusivity for elderly people and stroke patients.

REFERENCES

[1] Lu, L.L.M., Henn, P., O'Tuathaigh, C., Smith, S. (2024). Patient–healthcare provider communication and age-related hearing loss: A qualitative study of patients' perspectives. *Irish Journal of Medical Science* (1971-), 193(1): 277-284. <https://doi.org/10.1007/s11845-023-03432-4>

[2] Braun, M.N., Müller-Klein, A., Sopp, M.R., Michael, T.,

Link-Dorner, U., Lass-Hennemann, J. (2024). The human ability to interpret affective states in horses' body language: The role of emotion recognition ability and previous experience with horses. *Applied Animal Behaviour Science*, 271: 106171. <https://doi.org/10.1016/j.applanim.2024.106171>

[3] Asahioğlu, E.N., Göksun, T. (2023). The role of hand gestures in emotion communication: Do type and size of gestures matter? *Psychological Research*, 87(6): 1880-1898. <https://doi.org/10.1007/s00426-022-01774-9>

[4] Dare, T.O. (2023). Importance of non-verbal communication in building trust and rapport. *Sapientia Global Journal of Arts, Humanities and Development Studies*, 6(3): 37-47.

[5] Alsulaiman, M., Faisal, M., Mekhtiche, M., Bencherif, M., Alrayes, T., Muhammad, G., Alfakih, T. (2023). Facilitating the communication with deaf people: Building a largest Saudi sign language dataset. *Journal of King Saud University-Computer and Information Sciences*, 35(8): 101642. <https://doi.org/10.1016/j.jksuci.2023.101642>

[6] Shin, J., Musa Miah, A.S., Hasan, M.A.M., Hirooka, K., Suzuki, K., Lee, H.S., Jang, S.W. (2023). Korean sign language recognition using transformer-based deep neural network. *Applied Sciences*, 13(5): 3029. <https://doi.org/10.3390/app13053029>

[7] Johari, R.T., Ramli, R., Zulkoffli, Z., Saibani, N. (2023). MyWSL: Malaysian words sign language dataset. *Data in Brief*, 49: 109338. <https://doi.org/10.1016/j.dib.2023.109338>

[8] Cassim, M. R., Parry, J., Pantanowitz, A., Rubin, D.M. (2022). Design and construction of a cost-effective, portable sign language to speech translator. *Informatics in Medicine Unlocked*, 30: 100927. <https://doi.org/10.2139/ssrn.4431777>

[9] Zhou, Z., Tam, V.W., Lam, E.Y. (2022). A portable sign language collection and translation platform with smart watches using a BLSTM-based multi-feature framework. *Micromachines*, 13(2): 333. <https://doi.org/10.3390/mi13020333>

[10] Kasapbaşı, A., Elbushra, A.E.A., Omar, A.H., Yilmaz, A. (2022). DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals. *Computer Methods and Programs in Biomedicine Update*, 2: 100048. <https://doi.org/10.1016/j.cmpbup.2021.100048>

[11] Suharijito, Thiracitta, N., Gunawan, H. (2021). SIBI sign language recognition using convolutional neural network combined with transfer learning and non-trainable parameters. *Procedia Computer Science*, 179: 72-80. <https://doi.org/10.1016/j.procs.2020.12.011>

[12] Singh, D.K. (2021). 3D-CNN based dynamic gesture recognition for Indian sign language modeling. *Procedia Computer Science*, 189: 76-83. <https://doi.org/10.1016/j.procs.2021.05.071>

[13] Zhang, Y., Min, Y., Chen, X. (2021). Teaching Chinese sign language with a smartphone. *Virtual Reality & Intelligent Hardware*, 3(3): 248-260. <https://doi.org/10.1016/j.vrih.2021.05.004>

[14] Adithya, V., Rajesh, R. (2020). Hand gestures for emergency situations: A video dataset based on words from Indian sign language. *Data in Brief*, 31: 106016. <https://doi.org/10.1016/j.dib.2020.106016>

[15] Pacifici, I., Sernani, P., Falconelli, N., Tomassini, S.,

- Dragoni, A.F. (2020). A surface electromyography and inertial measurement unit dataset for the Italian Sign Language alphabet. *Data in Brief*, 33: 33195774. <https://doi.org/10.1016/j.dib.2020.106455>
- [16] Wadhawan, A., Kumar, P. (2020). Deep learning-based sign language recognition system for static signs. *Neural Computing and Applications*, 32(12): 7957-7968. <https://doi.org/10.1007/s00521-019-04691-y>
- [17] Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., AlKhalaf, R. (2019). ArASL: Arabic alphabets sign language dataset. *Data in Brief*, 23: 103777. <https://doi.org/10.1016/j.dib.2019.103777>
- [18] Ko, S.K., Kim, C.J., Jung, H., Cho, C. (2019). Neural sign language translation based on human keypoint estimation. *Applied Sciences*, 9(13): 2683. <https://doi.org/10.3390/app9132683>
- [19] Zhuang, C., Li, S., Ding, H. (2023). Instance segmentation based 6D pose estimation of industrial objects using point clouds for robotic bin-picking. *Robotics and Computer-Integrated Manufacturing*, 82: 102541. <https://doi.org/10.1016/j.rcim.2023.102541>
- [20] Chibane, J., Engelmann, F., Anh Tran, T., Pons-Moll, G. (2022). Box2mask: Weakly supervised 3D semantic instance segmentation using bounding boxes. In *European Conference on Computer Vision*, Tel Aviv, Israel, pp. 681-699. https://doi.org/10.1007/978-3-031-19821-2_39
- [21] Shandilya, S.K., Srivastav, A., Yemets, K., Datta, A., Nagar, A.K. (2023). YOLO-based segmented dataset for drone vs. bird detection for deep and machine learning algorithms. *Data in Brief*, 50: 109355. <https://doi.org/10.1016/j.dib.2023.109355>
- [22] Ulusoy, U., Eren, O., Demirhan, A. (2023). Development of an obstacle avoiding autonomous vehicle by using stereo depth estimation and artificial intelligence based semantic segmentation. *Engineering Applications of Artificial Intelligence*, 126: 106808. <https://doi.org/10.1016/j.engappai.2023.106808>
- [23] Zhang, J., Han, F., Han, D., Su, Z., Li, H., Zhao, W., Yang, J. (2023). Object measurement in real underwater environments using improved stereo matching with semantic segmentation. *Measurement*, 218: 113147. <https://doi.org/10.1016/j.measurement.2023.113147>
- [24] Srivastava, V., Mishra, S., Gupta, N. (2023). Automatic detection and categorization of road traffic signs using a knowledge-assisted method. *Procedia Computer Science*, 218: 1280-1287. <https://doi.org/10.1016/j.procs.2023.01.106>
- [25] Sanamdikar, S.T., Mayura, V.S., Rothe, J.P. (2023). Enhanced classification of diabetic retinopathy via vessel segmentation: A deep ensemble learning approach. *Ingenierie des Systemes d'Information*, 28(5): 1377. <https://doi.org/10.18280/isi.280526>
- [26] Shabrina, N.H., Lika, R.A., Indarti, S. (2023). Deep learning models for automatic identification of plant-parasitic nematode. *Artificial Intelligence in Agriculture*, 7: 1-12. <https://doi.org/10.1016/j.aiaa.2022.12.002>
- [27] Kimeu, J. M., Kisangiri, M., Mbelwa, H., Leo, J. (2024). Deep learning-based mobile application for the enhancement of pneumonia medical imaging analysis: A case-study of West-Meru Hospital. *Informatics in Medicine Unlocked*, 50: 101582. <https://doi.org/10.1016/j.imu.2024.101582>
- [28] Neamah, S.B., Karim, A.A. (2023). Real-time traffic monitoring system based on deep learning and YOLOv8. *Aro-the Scientific Journal of Koya University*, 11(2): 137-150. <https://doi.org/10.14500/aro.11327>
- [29] Casas, E., Ramos, L., Romero, C., Rivas-Echeverría, F. (2024). A comparative study of YOLOv5 and YOLOv8 for corrosion segmentation tasks in metal surfaces. *Array*, 22: 100351. <https://doi.org/10.1016/j.array.2024.100351>
- [30] Neamah, S.B. (2023). Real Time estimating size and speed of moving vehicles using deep learning. *Technology University's PhD dissertation*. <https://doi.org/10.13140/RG.2.2.32883.09760>
- [31] Syafaah, L., Faruq, A., Setyawan, N., Khair, M.I. (2024). Sick and dead chicken detection system based on YOLO algorithm. *Ingénierie des Systèmes d'Information*, 29(5): 1723-1729. <https://doi.org/10.18280/isi.290506>
- [32] Tamang, S., Sen, B., Pradhan, A., Sharma, K., Singh, V.K. (2023). Enhancing covid-19 safety: Exploring yolov8 object detection for accurate face mask classification. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2): 892-897.
- [33] Xiao, B., Nguyen, M., Yan, W.Q. (2024). Fruit ripeness identification using YOLOv8 model. *Multimedia Tools and Applications*, 83(9): 28039-28056. <https://doi.org/10.1007/s11042-023-16570-9>
- [34] Hasan, A.M., Diepeveen, D., Laga, H., Jones, M.G., Sohel, F. (2024). Object-level benchmark for deep learning-based detection and classification of weed species. *Crop Protection*, 177: 06561. <https://doi.org/10.2139/ssrn.4511105>
- [35] de Melo Lima, B.P., Borges, L.D.A.B., Hirose, E., Borges, D.L. (2024). A lightweight and enhanced model for detecting the Neotropical brown stink bug, *Euschistus heros* (Hemiptera: Pentatomidae) based on YOLOv8 for soybean fields. *Ecological Informatics*, 80: 102543. <https://doi.org/10.1016/j.ecoinf.2024.102543>
- [36] Solimani, F., Cardellicchio, A., Dimauro, G., Petrozza, A., Summerer, S., Cellini, F., Renò, V. (2024). Optimizing tomato plant phenotyping detection: Boosting YOLOv8 architecture to tackle data complexity. *Computers and Electronics in Agriculture*, 218: 108728. <https://doi.org/10.1016/j.compag.2024.108728>